IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 4, December 2023, pp. 1774~1783
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i4.pp1774-1783

Analysis of clustering and association using data mining technique for elderly health condition dataset

Rattanawadee Panthong[1], Thawin Wongkanthiya[2]
[1]School of Information and Communication Technology, University of Phayao, Phayao, Thailand
[2]Maeka Health Promotion Hospital, Maeka Sub-District, Muang District, Phayao, Thailand

Article Info

Article history:

Received Sep 30, 2022
Revised Jan 25, 2023
Accepted Jan 30, 2023

Keywords:

Association
Clustering
Data mining
Elderly
Health condition

ABSTRACT

Data survey on the elderly health condition in each year aimed to investigate the performance result on the elderly health care and to evaluate the elderly's health and health promotion. Thus, in analyzing the data, it mainly relied on the mining data technique for the evaluating health condition. This study presented the data analysis by clustering method. Then, the data was taken from each group to find the association rule. The analysis results showed that the elderly's health condition data could be classified into four different groups; cluster 1 (25%) were male elderly with high blood pressure and smoking cigarette, cluster 2 (25%) were female elderly with no the congenital disease but the result from the eye sight examination, it was found that they were long-sighted, cluster 3 (24%) were female elderly with no the congenital disease but having the insomnia and osteoarthritis and cluster 4 (26%) were female elderly with high blood pressure and diabetes. It also indicated that each group had the rule showing the correlation between the data in each group having the minimum value of confidence at 0.8 and the minimum value of support not less than 0.5.

Corresponding Author:

Rattanawadee Panthong
School of Information and Communication Technology, University of Phayao
19 moo2, Phaholyothin Road, Muang, Phayao, Thailand
Email: rattanawadee.pa@up.ac.th

1. INTRODUCTION

Thailand is entering an aging society increasingly and rapidly. In 2022, it indicated that number of the elderly, aged over 60 years old is 18.30% of Thailand's population and expected that in the next twenty years, Thailand will be a super aging society [1], [2]. Also, it is likely that the elderly will stay alone more and more which will have an effect on the elderly's health care and health promotion, essential for carrying their daily lives [1]–[3].

In phayao province, it was found that in 2021, numbers of the elderly aged over 60 years old in an area of Maeka Health Promotion Hospital were 1,363 or 23.26% of the total population which is absolutely an aging society. This hospital mainly focused on the health problems and readiness in health promotion to prevent the elderly's diseases such as diabetes, blood pressure, heart attack, renal ailment, stoke and more vegetative patients as a result of their heath conditions and their diseases [4]. Moreover, there was a survey on the alert of the elderly health condition in each year in order to survey the problem of the elderly's health condition follow the results of the elderly's health care treatment from different ailments involved with the elderly's doing their daily routine through Barthel activities of daily living (ADL) [5].

In planning on the elderly health care, it is essential to use the data base to solve the problem for a long time in order to find the guideline to cooperate with other sections in the community, in evaluating the elderly's health condition relies on the important data. In these days, each organization has more data leading

to the difficulty in management. The data analytic rely on technology of data mining and the machine learning to help increase the effectiveness of data analysis.

Thus, the researchers applied the data from survey on the elderly's health condition through a data mining technique. In this study, the data mining technique was applied to analyze the data by clustering with K-means algorithm and association rule with frequent pattern-Growth (FP) approach. These methods are an unsupervised learning.

The objectives of this study aimed to analyze clustering of data survey on the elderly health condition and to find association rule of each cluster. The expected outcomes from analysis of clustering and the association rule were the village health care volunteers and the related sectors who could take the data analysis to be the model and the guideline for developing the elderly's health care better and more suitable than before. Furthermore, these data could be applied to plan the patient's health care promotion in each area.

2. RESEARCH BACKGROUND

2.1. Barthel activities of daily living (ADL)

ADL is the ability measurement of the elderly in leading their daily lives. Level of ADL evaluation has 4 categories including: score 0-4 very low initial score, total dependence, score 5-8 low initial score, severe dependence, score 9-11 intermediate initial score, moderately severe dependence, score 12-20 intermediate high, mildly severe dependence, consideration of discharging home [3]. It is a set of 10 question items about the ability to help oneself in doing routine activities [5] as the following.

- Eating when food is served before face.
- Washing face, combing hair, brushing teeth and rasing beard and mustache within last 24-48 hours.
- Get up from a bed to sit on a chair.
- Using a toilet or a bathroom
- Moving in a room or a house.
- Putting clothes
- Step up and down the stairs
- Taking bath
- Holding a bowel movement within last week
- Holding urinating within last week.

2.2. Data mining

Data mining is data analysis from a large of datasets. Data analytics with data mining technique to find hidden patterns and rules or relationships. This method aims to acquire new knowledge that can be used to make decisions in various fields [6]. Data mining is used to solve problem in business, marketing, education or healthcare. The data mining has several techniques such as clustering, association rule, classification, and regression [7].

2.3. Clustering in machine learning

Clustering is also called segmentation analysis. It is unsupervised learning technique to machine learning [8]. This method is divided into at least two subgroups. The concept discriminant analysis is based on the characteristics of the most similar data in the same group and the characteristics of each group of data which are mostly the most different [9]. The examples of clustering are grouping customers by shopping behavior, grouping tourists based on their behavior or grouping patients according to symptoms or severity of disease [10], [11].

2.4. K-means

K-means is a simple unsupervised machine learning algorithm. It is non-hierarchical cluster analysis or partition. The k-mean technique is a way to classify the cluster into K cluster as determined by the user [8]. It is a popular method used to classify the similar data into a partitional clustering and then measuring the distance between each data and the centroid [12], [13].

Steps of clustering by K-means:

- Determine the center of each cluster by sampling as the amount needed.
- Calculate the distance between each data and the centroid of each cluster via Euclidean distant method.
- Calculate the mean value of data each cluster and use it to be the centre each cluster replaced the old one.
- Step 2 and 3 are replaced until every data being in the old cluster.

Elbow method is a technique used for measuring the error of the total distance between the data and the centroid or within groups sum of squares (WGSS) in order to choose the K value which is appropriate or

find optimal cluster number where the amount of the appropriate clusters is the elbow point [14], [15]. The optimal of K value with the elbow method is calculated,

$$\text{WCSS}= \sum_{C_k}^{C_n}(\sum_{d_i in C_i}^{d_m} distance(d_i, C_k)^2) \quad (1)$$

C is the cluster centroids and D are the data point in each cluster.

2.5. The association rules

Finding the association rule is an important technique of data mining and can be applied to different tasks. Its working principle of this method is to find the data association from the large data existing to be applied to analyze or predict different situations or analyze the consumer's merchandise purchase which is known as market basket analysis evaluated form the data in table which is gathered [16]–[18]. The analysis result gained will be the answer of question which is the analysis using the association rule to find the data correlation [19]. In choosing the rule having an importance will be measured from the measurement value as

$$\text{Support } (X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (2)$$

$$\text{Confidence } (X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (3)$$

2.6. Frequent pattern (FP) growth algorithm in data mining

FP-growth is a way to look for a model of data often occurring which is an algorithm in data association without creating candidate itemset [20], [21]. Its stage's function is to use the FP-tree structure stating from reading data in order to create the tree and them detect the in-depth data to find the data appearing together frequently. This method will read the data from the database only two times in order to reduce time consumption in processing more quickly [22], [23].

2.7. Related work

Cheng *et al.* [24] proposed a method of clustering analysis and association rule technology in cross-marketing. In this work, the PrefixSpan algorithm was improved. The experimental results showed that the PrefixSpan algorithm could reduce the time consumption in constructing the projection database and the influence of the increase of support on the algorithm efficiency and could improve the accuracy and efficiency of algorithm.

Shahin *et al.* [11] proposed method of the k-modes as the main clustering method and the association rule with Apriori algorithm in an accident data with 576 instances. The grouping using the k-modes algorithm to discover different 4 clusters and then the Apriori method to find correlation od each cluster. These methods can help define the segment circumstances associated with an accident.

In addition, Yotsawat and Srivihok [25] presented the clustering of tourist behavior using two steps (Som and K-means) algorithms. The Som algorithm is applied to find the optimum number of groups. The research results received 11 different clusters and the distinct features among groups can be helpful for tourism organizations in order to determine the market planning or to design package tour.

3. RESEARCH METHOD

3.1. Dataset

The dataset used in this research was gained from Maeka Health Promotion Hospital, Maeka Sub-District, Muang District, Phayao Province in Thailand. The data of the elderly's health condition was collected from January 2022 to March 2022. Details of the dataset feature were shown in Table 1.

Table 1. Showing the details of the data feature used in the experiment

Dataset	Number of features	Number of instances	Type of features
Meaka_ADL	41	1,194	Nominal, Numeric

3.2. Proposed method

The analysis of clustering and association rules uses the data mining technique for elderly health condition dataset. The K-means algorithm is used to divide groups of a dataset. The correlation of each cluster

was found using FP-growth method. The framework for the analysis of clustering and association rules using data mining technique is illustrated in Figure 1.

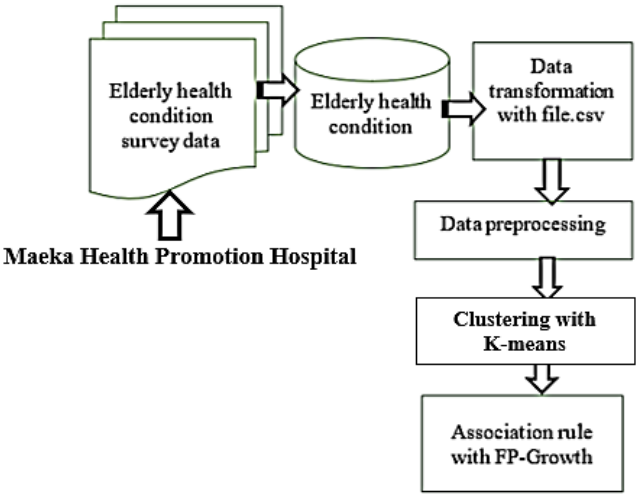

Figure 1. Framework of clustering and association analysis with data mining technique

The steps of this research were as followed,

- Collecting the data from the survey questionnaire was kept in database.
- Changing the data into csv file format which was easy to apply.
- Conducting the preprocessing data in order to make the data accurate.

Managing the missing values,

- Listwise deletion or complete analysis was used in the case that data loss many features but not more than 5% of the whole data by erasing the instance having lost data [26].
- Imputation technique with mean imputation (MI) is a method of calculating the mean of the independent variable (x) from the appearing dataset in order to replace the missing value of the dependent variable (y), This approach is an easy way in managing the lost data [27].

In this step, missing values estimation and discretization were used before clustering approaches. The missing nominal features were replaced by unknown value [28]. In addition, the numerical attributes were discretized by referencing value in the laboratory or user-defined values.

- Clustering data in segmentation the similar data in the same cluster, in this study, K-Means algorithm was used. In this process, the Elbow method is applied to calculate the appropriate K value as shown in Figure 2.
- Finding the data association in each cluster by using FP-growth algorithm was to find association occurring together in data. In this work, the minimum support value was 0.6 and minimum confidence was 0.8. Furthermore, the present study used WEKA version 3.8.6 [29]. It was used as a tool for data preparation, clustering, and association rules mining.

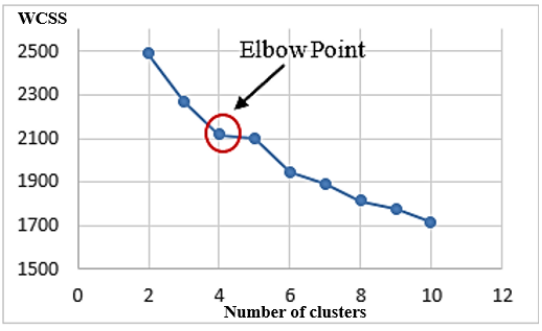

Figure 2. Showing the selection of optimal K clusters with the elbow method

4. RESULTS AND DISCUSSION

The results of clustering analysis for elderly health situation dataset indicated that the elderly segments four different clusters. In Figure 3 demonstrated percentage of elderly amount from each cluster. Furthermore, the feature of each cluster could be explained.

i) Cluster 1 elderly cluster with blood pressure (25%): Male elderly most of them did not stay alone, having spouse's care, a government pensioner and having blood pressure, the examination result, it was found that the mean value of blood pressure was high (140-159/90-99), Smoking and moderate movement activity were having but irregular (less than 3 times per week), in the present having good teeth at least 20 teeth, eye sight examination was normal, no problem of sleeping and no knee pain or osteoarthritis.

ii) Cluster 2 elderly cluster with abnormal eye sight, long sighted (25%): Female elderly most of them did not stay alone, they stayed with their spouse and cousins, unemployment, no chronic disease, no smoking, blood pressure examination result, it showed that the average was stared to be high (120-139/80-89), moderate movement activity were having regularly (at least 3 times a week), now having good teeth at least 20 teeth, no problem of sleeping and no knee pain or osteoarthritis and eye-sight examination; it was found that they had eye sight problem which was long-sighted.

iii) Cluster 3 elderly cluster with the problems of sleeping and having knee pain or osteoarthritis (24%): Female elderly most of them did not stay alone, stayed with their spouse and cousins take care, earn their living by farming, no chronic disease and no smoking, blood pressure examination result, it showed that the average was stared to be high (120-139/80-89), moderate movement activity were having regularly (at least 3 times a week), in the present having good teeth 20 teeth, the result of eye sight examine, it was found that it was normal, but having problem of the insomnia and knee pain or osteoarthritis.

iv) Cluster 4 elderly cluster with chronic diseases; diabetes and blood pressure (26%): Female elderly most of them did not stay alone, they stayed with their spouse and cousins, no working, chronic diseases; diabetes and high blood pressure, blood pressure examination, it showed that the blood pressure was high (140-159/90-99) but not smoking, moderate movement activity, they couldn't do it, Currently, they have good teeth no more than 20 teeth, eye sight examination was normal, no problem of sleeping and no knee pain or osteoarthritis.

The data distribution according to their similar feature was shown in Table 2 as shown in Appendix and each cluster had the distant measurement between the cluster by a means of Euclidean Distance from the center of each cluster as in Table 3. Furthermore, the result of finding the association rule by using FP-growth technique when data clustering of the elderly having the similar features was done, the association rule occurring together significantly of each feature through FP-growth was calculated. In this experiment, the determined value of minimum support is 0.5 and minimum confidence is 0.8. For Table 4 illustrated the example of rules gained from each cluster.

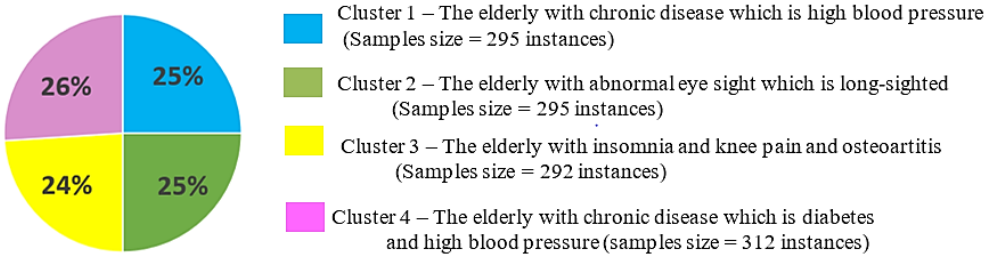

Figure 3. Percentage of elderly amount from each group

Table 2. The significant characteristics of elderly health situation in 4 clusters (*continue…*)

Feature	Clusters				Feature	Clusters			
	1	2	3	4		1	2	3	4
1. Gender					2. Situ1 (Staying alone)				
1=male	295	0	165	82	1=Yes	20	35	28	26
2=Female	0	295	127	230	0=No	275	260	264	286
3. Situ2 (Having spouse care)					4. Situ3 (Immediate family care)				
1=Yes	284	26	0	139	1=Yes	105	183	20	191
0=No	11	269	292	173	0=No	190	112	272	121

Table 2. The significant characteristics of elderly health situation in 4 cluster

5. Situ4 (Having cousin care)					6. Situ5 (Relative care)				
1=Yes	5	15	23	27	1=Yes	5	9	13	28
0=No	290	280	269	285	0=No	290	286	279	284
7. Situ6 (No having cousin care care)					8. Cdis (Having chronic disease)				
1=Yes	1	2	0	2	1=Yes	147	145	122	311
0=No	294	293	292	310	0=No	146	148	169	1
					Unknown	2	2	1	0
9. Cdis1 (Diabetes disease)					10. Cdis2 (High blood pressure disease)				
1=Yes	27	40	14	102	1=Yes	46	0	30	312
0=No	268	255	278	210	0=No	249	295	262	0
11. Cdis3 (Cholesterol disease)					12. Cdis4 (Cardiovascular disease)				
1=Yes	6	14	17	123	1=Yes	7	8	7	19
0=No	289	281	275	189	0=No	288	287	285	293
13. Cdis5 (Cerebrovascular disease)					14. Cdis6 (Kidney failure disease)				
1=Yes	1	2	6	7	1=Yes	4	2	3	8
0=No	294	293	286	305	0=No	291	293	289	304
15. Cdis7 (Asthma disease)					16. Cdis8 (Emphysema disease)				
1=Yes	5	4	5	4	1=Yes	3	1	4	2
0=No	290	291	287	308	0=No	292	294	288	310
17. Cdis9 (cancer disease)					18. Cdis10 (anemia disease /thalassemia)				
1=Yes	1	1	3	5	1=Yes	0	2	1	3
0=No	294	294	289	307	0=No	295	293	191	309
19. Cdis11 (schizophrenia disease)					20. Cdis12 (epilepsy disease)				
1=Yes	1	4	5	4	1=Yes	0	0	0	3
0=No	294	291	287	308	0=No	295	295	292	309
21. Cdis13 (gout disease)					22. Cdis14 (Knee osteoarthritis disease)				
1=Yes	4	3	7	12	1=Yes	11	13	20	48
0=No	291	292	285	300	0=No	284	282	272	264
23. BP1 (Blood pressure)					24. Occ (Occupation)				
1= normal (not more than 120/80)	62	88	75	62	1=Agriculture	83	35	81	53
2=starting high (120-139/80-89)	150	149	156	160	2=General employee	69	41	52	39
3=high (140-159/90-99)	58	42	43	58	3=Merchant/ Personal business	24	40	45	23
4=very high (>160/100)	23	11	14	31	4=Animal keeper	8	4	5	4
0=no information	2	5	4	1	5=Government pensioner	20	17	10	10
25. Mact (Moderate movement activity)					6=Unemployment	86	153	81	166
1=Having regularly at least 3 times a week	176	166	172	150	7=Others	3	4	15	17
2=Having irregularly less than 3 times a week	93	106	103	124	0=Unknown	2	1	3	0
3=No having due to disability	25	22	17	36	26. Smoking				
0=Unknown	1	1	0	2	Never smoking	112	259	204	230
27. Results_eye (the result of eye sight examine)					Ever smoking but not smoking now	104	22	38	60
1=short-sighed eye	28	31	21	25	Smoking	79	14	50	22
2=long-sighed eye	47	52	64	62	28. sleep_problem (Details of sleep problem diagnosis)				
3=cataract	13	9	11	19	1=insomnia	34	65	86	105
4=glaucoma	2	1	1	4	2=oversleeping	4	10	3	10
5=pterygium	2	4	7	8	3=snoring	20	6	11	12
6=macular degeneration	5	6	6	9	4=waking up at night from dreaming	2	0	0	2
0=Unknown	198	192	182	185	0=Unknown	235	214	192	183
29. Assess_sleep (Sleep problem diagnosis)					30. Teeth (The present having good teeth at least 20 teeth)				
1=Yes	59	80	103	126	1=Yes	236	229	241	236
2=No	234	214	188	183	2=No	57	60	50	72
Unknown	2	1	1	3	Unknown	2	6	1	4
31. Assess_knee (Knee osteoarthritis diagnosis)					32. ADL1 (Feeding)				
1=Yes	112	157	149	204	0= unable	1	1	2	5
2=No	182	138	143	108	1= needs help cutting, and spreading butter.	24	19	13	31
Unknown	1	0	0	0	2=independent (food provided within reach	270	275	277	276

Table 2. The significant characteristics of elderly health situation in 4 cluster

33. ADL2 (Grooming)					34. ADL3 (Transfer)				
0 = needs help with personal care	4	3	2	10	0 = unable – no sitting balance	1	0	0	3
1 = independent face/hair/teeth/shaving (implements provided)	291	292	290	302	1 = major help (one or two people, physical), can sit	1	1	4	16
35. ADL4 (Toilet use)					2 = minor help (verbal or physical)	13	13	6	25
0 = dependent	2	2	1	9	3 = independent	280	277	282	268
1 = needs some help, but can do something alone	9	11	5	22	36. ADL5 (Mobility)				
2 = independent (on and off, dressing, wiping)	284	282	286	281	0 = immobile	2	2	1	5
37. ADL6 (Dressing)					1=wheelchair independent, and including corners.	3	1	2	16
0 = dependent	2	2	1	8	2 = walks with help of one person (verbal or physical)	11	13	8	22
1 = needs help, but can do about half unaided	4	6	2	14	3 = independent	279	279	281	269
2 = independent (including buttons, zips, and laces)	298	286	289	290	38. ADL7 (Stairs)				
39. ADL8 (Bathing)					0 = unable	3	5	4	18
0 = dependent	5	7	2	16	1 = needs help (verbal, physical, carrying aid)	7	13	7	19
1 = independent (or in shower)	290	288	290	296	2 = independent up and down	285	277	281	275
40. ADL9 (Bowels)					41. ADL10 (Bladder)				
0 = Incontinent (or needs to be given enema)	3	2	2	6	0 = incontinent, or catheterized and unable to manage	3	2	1	7
1 = occasional accident (once/week)	15	24	6	35	1 = occasional accident (max. once per 24 hours)	16	30	14	47
2 = continent	277	289	284	271	2 = continent (for over 7 days)	276	263	277	258

Table 3. Centroid of each cluster

Cluster	Cluster centroids			
	1	2	3	4
1		1.70	1.30	2.02
2	1.70		1.63	1.47
3	1.30	1.63		1.87
4	2.02	1.47	1.87	
Number of instances	295	295	292	312

Table 4. Illustrated the example of rule gained from each cluster with Minsup=40%-50%, Minconf=80%-95%

No.	Premises	Conclusion
R_{11}	Teeth = 1, Cdis=1	Situ6=0, ADL7 = 2, ADL4 = 2, ADL3 = 3
R_{12}	Cdis=1	sleep_detail = 0
R_{13}	ADL7 = 2, Cdis=1	ADL6 = 2, Teeth = 1, Smoking = 3
R_{21}	ADL4 = 2, Cdis=0	Smoking = 1
R_{22}	ADL5 = 3, ADL3 = 3, ADL7 = 2, ADL9 = 2, Cdis=0	Situ4=1
R_{23}	ADL6 = 2, Cdis=0	Situ4, ADL9 = 2, Situ2 = 1
R_{31}	ADL7 = 2, Assess_knee=1	ADL6 = 2, ADL4 = 2, ADL3 = 3, ADL5 = 3, Teeth = 1
R_{32}	ADL9 = 2, Assess_knee=1	ADL6 = 2, ADL3 = 3, Teeth = 1
R_{33}	BP1 = 2	ADL6 = 2, ADL4 = 2, ADL9 = 2, ADL7 = 2, ADL10 = 2
R_{41}	ADL6 = 2, Smoking = 1	Cdis1=1
R_{42}	ADL4 = 2, Teeth = 0	Cdis1=1
R_{43}	ADL6 = 2, ADL1 = 2, ADL9 = 2, Teeth = 0	Cdis1=1

5. DISCUSSION

In overall results, it revealed that data clustering via K-means algorithm could give the different clustering features in 4 clusters; cluster1-the elderly cluster with chronic disease which was high blood pressure (%), cluster 2-the elderly cluster with the eye sight problem which was long-sighed (%), cluster 3-the elderly cluster with insomnia and knee pain (%) and cluster 4-the elderly cluster with diabetes and high blood pressure (%). The result of this research which was different from the previous study [25] was that this study applied K-means technique and FP-growth method in elderly health condition dataset clustering and then the data from each cluster was sought for its association in specific feature of each group and it was refined and analyzed by the experts. The benefits obtained from this rule can be applied to plan caring the elderly's health in the responsible area of Meaka Heath Promotion Hospital and can be used to promote activities the elderly's health

care. The rule shown out will be refined from the experts. It still found that there was an interesting rule which were demonstrated in Table 5.

Table 5. The implementation of clustering and association rule for elderly health care (Continue…)

Knowledge found from association rules	Activities which should be suggest for elderly
Most of rules in cluster 1 are have chronic disease, smoking and the present having good teeth at least 20 teeth.	Guideline of planning in taking care of the elderly, group 1 – Manage the service system concerning caring the chronic patients of the public health unit to cover the target and take care continuously by a specialist doctor and a practical nurse. – Provide an advising clinic in changing the patient's behaviors. – Set a plan in changing the chronic patient's behaviors. Activity in promoting the elderly health care, group 1 – Giving knowledge in changing the patient's behaviors; food, exercise, emotion and drinking or smoking. – Having a yearly check up to look for complications in patients. – Develop the model person who can change his or her behavior to give his/her knowledge during the treatment period. – Visit the patient's home, examine and tackle each case.
For cluster 2, if "Cdis=0" and "ADL4=2" then "Smoking = 1"	Guideline of planning in taking care of the elderly, group 2 Transfer the patient to the specialist or the hospital network to be examined and treated Set up the service-advising system in a primary care unit. Set up a project or plan on optic components and diseases. Activity in promoting the elderly health care, group 2 Give the advice concerning optic components and diseases in a primary care unit. Check the eye sight first before sending to the network hospitals.
For cluster 3, if "Assess_knee=1" and "ADL4=2" then "Smoking = 1"	Guideline of planning in taking care of the elderly, group 3 – Have an advising clinic for insomnia, osteoarthritis integrating with School of Allied Health Sciences, University of Phayao. – Have a project or plan concerning osteoarthritis and neurosis from the sub district health fund. – Integrate the caring the activities of the patients of knee problems with School of Allied Health Sciences, University of Phayao. Activity in promoting the elderly health care, group 3 – Assess the knee osteoarthritis in the elderly group. – Promote exercising, losing weight and controlling diet – Gather to do activities in the elderly school of the sub district's primary health care hospital. – Have the checking up and caring activities by faculties and students from School of Allied Health Sciences, University of Phayao. – Visit the patient's home, examine and tackle each case
Most of rules in cluster 4 are have good teeth no more than 20 teeth and have chronic disease.	Guideline of planning in taking care of the elderly, group 4 – Manage the service system concerning caring the chronic patients of the public health unit to cover the target and take care continuously by a specialist doctor and a practical nurse. – Provide an advising clinic in changing the patient's behaviors – Set a plan in changing the chronic patient's behaviors Activity in promoting the elderly health care, group 4 – Develop the model person who can change his or her behavior to give his/her knowledge during the treatment period. – Having a yearly check up to look for complications in patients. – Giving knowledge in changing the patient's behaviors; food, exercise, emotion and drinking or smoking. – Visit the patient's home, examine and tackle each case – Gather to do activities in the elderly school of the sub district's primary health care hospital.

6. CONCLUSION

This research presented the clustering method via K-means algorithm and sought for the association rule through FP-growth approach. The dataset on the elderly health situation of Meaka Health Promotion Hospital collected from 1,194 instances was divided into 4 clusters. The clustering outputs indicated that health condition of the elderly consisted of various groups with different characteristics. FP-growth method was applied in order to discover the relationship among the features of elderly in each group. The association rules were evaluated by experts. The result from data analysis could be used to plan the elderly health care in communication and could be prototype model for developing or creating activities in the elderly's health care condition. In addition, from the data analysis of the obtained rules, it indicated that the association rules had the interesting rule and the majority of the acquired rules are the outstanding feature of each cluster. However,

some rules did not show the association indicating the distinct feature of the feature of each cluster. Due to the data imbalance, the preparing data process should be adapted in terms of the balance of the dataset before analyzing the data in order to gain the rule which is reliable and having more association. For the future work, in data analysis, other mining data might be applied on refining more knowledge.

ACKNOWLEDGEMENTS

This research was supported by School of Information and Communication Technology, University of Phayao, Thailand (Grant No. RMICT65013).

REFERENCES

[1] K. Kantachote and N. Wiroonsri, "Do elderly want to work? Modeling elderly's decision to fight aging Thailand," *Quality and Quantity*, 2022, doi: 10.1007/s11135-022-01366-0.
[2] S. Jitapunkul and S. Wivatvanit, "National policies and programs for the aging population in Thailand," *Ageing International*, vol. 33, no. 1–4, pp. 62–74, 2008, doi: 10.1007/s12126-009-9027-6.
[3] S. Suriyanrattakorn and C. L. Chang, "Long-term care (LTC) policy in Thailand on the homebound and bedridden elderly happiness," *Health Policy OPEN*, vol. 2, 2021, doi: 10.1016/j.hpopen.2020.100026.
[4] M. Rugbumrung *et al.*, "The quality of life of the ageing in Phayao Province," *Journal of Positive School Psychology*, vol. 6, no. 8, pp. 1474–1480, 2022.
[5] C. Collin, D. T. Wade, S. Davies, and V. Horne, "The barthel ADL index: A reliability study," *Disability and Rehabilitation*, vol. 10, no. 2, pp. 61–63, 1988, doi: 10.3109/09638288809164103.
[6] N. Jothi, N. A. Rashid, and W. Husain, "Data mining in healthcare - A review," *Procedia Computer Science*, vol. 72, pp. 306–313, 2015, doi: 10.1016/j.procs.2015.12.145.
[7] M. J. Hamid Mughal, "Data mining: Web data mining techniques, tools and algorithms: An overview," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, pp. 208–215, 2018, doi: 10.14569/IJACSA.2018.090630.
[8] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
[9] M. Dave and H. Gianey, "Different clustering algorithms for big data analytics: A review," *Proceedings of the 5th International Conference on System Modeling and Advancement in Research Trends, SMART 2016*, pp. 328–333, 2017, doi: 10.1109/SYSMART.2016.7894544.
[10] H. D. Meng, Y. C. Song, F. Y. Song, and H. T. Shen, "Application research of cluster analysis and association analysis," *2nd International Conference on Software Engineering and Data Mining, SEDM 2010*, pp. 597–602, 2010.
[11] M. Shahin *et al.*, "Cluster-based association rule mining for an intersection accident dataset," *2021 International Conference on Computing, Electronic and Electrical Engineering, ICE Cube 2021 - Proceedings*, 2021, doi: 10.1109/ICECube53880.2021.9628206.
[12] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration k-means clustering method and elbow method for identification of the best customer profile cluster," *IOP Conference Series: Materials Science and Engineering*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.
[13] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics (Switzerland)*, vol. 9, no. 8, pp. 1–12, 2020, doi: 10.3390/electronics9081295.
[14] T. M. Kodinariya and P. R. Makwana, "Review on determining of cluster in K-means," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, no. 6, pp. 90–95, 2013, [Online]. Available: https://www.researchgate.net/publication/313554124.
[15] H. Humaira and R. Rasyidah, "Determining the appropiate cluster number using elbow method for k-means algorithm," 2020, doi: 10.4108/eai.24-1-2018.2292388.
[16] A. Saxena and V. Rajpoot, "A comparative analysis of association rule mining algorithms," *IOP Conference Series: Materials Science and Engineering*, vol. 1099, no. 1, p. 012032, 2021, doi: 10.1088/1757-899x/1099/1/012032.
[17] N. H. Miswan, I. M. Sulaiman, C. S. Chan, and C. G. Ng, "Association rules mining for hospital readmission: A case study," *Mathematics*, vol. 9, no. 21, 2021, doi: 10.3390/math9212706.
[18] D. J. Prajapati, S. Garg, and N. C. Chauhan, "Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment," *Future Computing and Informatics Journal*, vol. 2, no. 1, pp. 19–30, 2017, doi: 10.1016/j.fcij.2017.04.003.
[19] U. Y.Bhat and P. A. Patel, "A recent overview: Rare association rule mining," *International Journal of Computer Applications*, vol. 107, no. 18, pp. 1–4, 2014, doi: 10.5120/18848-9893.
[20] S. Kanhere, A. Sahni, P. Stynes, and P. Pathak, "Clustering based approach to enhance association rule mining," *Conference of Open Innovation Association, FRUCT*, vol. 2021-January, 2021, doi: 10.23919/FRUCT50888.2021.9347577.
[21] S. Sidhu, U. Kumar Meena, A. Nawani, H. Gupta, and N. Thakur, "FP growth algorithm implementation," *International Journal of Computer Applications*, vol. 93, no. 8, pp. 6–10, 2014, doi: 10.5120/16233-5613.
[22] A. P. U. Siahaan, M. Mesran, A. H. Lubis, and A. Ikhwan, "Association rules analysis on FP-growth method in predicting sales," *International Journal of Recent Trends in Engineering and Research*, vol. 3, no. 10, pp. 58–65, 2017, doi: 10.23883/ijrter.2017.3453.dhcoa.
[23] Y. Zeng, S. Yin, J. Liu, and M. Zhang, "Research of improved FP-growth algorithm in association rules mining," *Scientific Programming*, vol. 2015, 2015, doi: 10.1155/2015/910281.
[24] Y. Cheng, M. Cheng, T. Pang, and S. Liu, "Using clustering analysis and association rule technology in cross-marketing," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/9979874.
[25] W. Yotsawat and A. Srivihok, "Rules mining based on clustering of inbound tourists in Thailand," *Lecture Notes in Electrical Engineering*, vol. 315, pp. 693–705, 2015, doi: 10.1007/978-3-319-07674-4_65.
[26] W. Junthopas and C. Wongoutong, "Comparison of listwise deletion and imputation methods for handling a single missing response value in a central composite design," *Thailand Statistician*, vol. 20, no. 3, pp. 545–561, 2022.

[27] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Frontiers in Energy Research*, vol. 9, 2021, doi: 10.3389/fenrg.2021.652801.
[28] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, "Big data preprocessing: Methods and prospects," *Big Data Analytics*, vol. 1, no. 1, 2016, doi: 10.1186/s41044-016-0014-0.
[29] Z. Markov and I. Russell, "An introduction to the WEKA data mining system," *ITiCSE06 - Proceedings of the 11th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education*, vol. 2006, pp. 367–368, 2006, doi: 10.1145/1140123.1140127.

BIOGRAPHIES OF AUTHORS

Rattanawadee Panthong received the Ph.D. (Computer Science) from Kasetsart University, Thailand in 2021, with the Dissertation "Hybrid Feature Selection based on Class Dependent Technique for Liver Cancer Classification Model of Central Region in Thailand". She is currently an instruction at School of Information and Communication Technology in University of Phayao, Thailand. In addition, she is serving as Assistant Dean in School of Information and Communication Technology (2019–present). Her research interests are in Machine Leaning, Data Mining, Data analytics and Data warehouse. She can be contacted at email: rattanawadee.pa@up.ac.th.

Thawin Wongkanthiya received the M.P.H. (Public Health Administration) from, Naresuan University, Thailand since 2006. In addition, He is serving as director of Maeka Subdistrict Health Promoting Hospital. Phayao, Thailland (2020–present). Him research interests are in primary care unit, Long-term Elderly Health Care System (LTC) and elderly health problems. He can be contacted at email: thawin04@gmail.com.